

Données textuelles : sommaire

Nous entendons par données textuelles des données permettant la représentation de textes peu structurés, tels qu'ils sont produits par les logiciels de traitement de texte ou les éditeurs de texte. Pour l'archivage de données textuelles, il est traditionnellement proposé de convertir les formats naturels de tels logiciels dans des formats plus appropriés aux formats d'archivage, p. ex. [PDF/A-1](#) ou [PDF/A-2](#). Lors du passage de formats de fichiers ouverts, fondés sur XML, dans des logiciels de traitement de texte, la question se pose de savoir si ceux-ci pourraient également se prêter aux formats d'archivage.

Il faut faire attention à la distinction fondamentale entre les formats basés sur les pages et ceux non basés sur les pages. Les formats basés sur les pages comme les [PDF](#) et les [PDF/A](#) sont optimaux pour l'impression, car le saut de page et la mise en page sont déterminés. Pour les formats non basés sur les pages, comme tous les formats de traitement de texte, le texte lui-même est la préoccupation centrale; le saut de page est seulement déterminé lors de l'impression (ou lors de la conversion dans un format basé sur les pages). Les données de présentation (p. ex. PowerPoint) font également partie des données textuelles et constituent un sous-type. Ces formats ne sont cependant pas traités de manière spécifique dans la présente version du Cfa.

Formats analysés

- [TXT](#)
- [PDF](#)
- [PDF/A-1](#)
- [PDF/A-2](#)
- [PDF/A-3](#) (n'est pas un format bien approprié pour l'archivage à long terme)
- [ODF](#)
- [OOXML](#)

Recommandation

Les données textuelles doivent, si possible, déjà être converties dans un format PDF (de préférence [PDF/A-2](#)) pendant la phase active et archivées sous cette forme.

La plupart du temps, la conversion ultérieure de fichiers PDF au format PDF/A ne pose pas de problème.

TXT

Informations générales

Titre	Texte brut; Plain Text; Text only
Catégorie	Formats textuels
Abréviation	TXT
Extension de fichier	.txt, .ans, .asc
Mime Type	text/plain
Pronom PUID	x-fmt/14; x-fmt/15; x-fmt/16; x-fmt/21; x-fmt/22; x-fmt/111; x-fmt/130; x-fmt/282; x-fmt/283
Version	Aucune; voir toutefois les informations sur le codage des caractères .

Description

Le format textuel peut être considéré comme le format le plus stable de l'informatique; le codage ASCII est déjà connu et utilisé depuis des décennies; les codages plus récents, comme la famille ISO-8859 ainsi que les divers codages UNICODE sont aussi rétrocompatibles.

Evaluation

Ouverture du format : 4

Les codages de caractères les plus usités sont les standards ISO ou IETF et sont donc ouverts.

Licence libre : 3

Il n'existe pas de licences de formats textuels ou de codage.

Diffusion : 4

Ce format est l'un des plus répandus de l'informatique mais il n'est que peu utilisé dans l'administration.

Fonctionnalités : 1

Les fichiers en texte brut ne permettent pratiquement pas (éventuellement sous forme rudimentaire) de représenter la mise en page (layout) et les formatages. Une migration des fichiers issus des programmes de traitement de texte en texte brut entraîne la perte presque complète du formatage, ce qui peut être inacceptable selon l'importance de ce dernier. En outre, la majeure partie des fichiers créés à partir de programmes de traitement de texte contiennent non seulement du texte, mais aussi des images, des graphiques, des tableaux et d'autres composants semblables, qui ne peuvent pas être pris en charge lors de la compression.

Implémentation : 4

Il existe un grand nombre d'éditeurs de texte parmi lesquels figurent également des produits open source

Densité de mémorisation : 4

La densité de mémorisation est élevée et peut être encore accrue par l'usage de la compression sans pertes.

Vérifiabilité : 2

La reconnaissance est principalement possible au moyen de l'extension. Il n'existe aucun validateur. Il est toutefois possible d'identifier le codage avec une analyse de plausibilité.

Bonnes pratiques : 4

Le texte brut est diffusé en tant que format d'archivage recommandé et utilisé.

Perspectives : 1

Le potentiel de développement du texte brut a pratiquement atteint ses limites.

Classe de formats : A

Il s'agit de l'un des plus anciens formats de l'informatique.

Conclusion

Le texte brut répond au mieux à la plupart des exigences de l'archivage (à l'exception près, non négligeable, de la préservation de la mise en page) et n'est donc pas contesté comme format d'archivage. L'administration n'utilise cependant presque pas les fichiers texte brut. Les fichiers XML, HTML et CSV constituent une exception car il s'agit en fait de pures données textuelles contenant aussi des informations sur la structure et/ou la mise en page, tout en étant destinées à une application spécifique (sites internet, contenu d'une base de données). En outre, certains systèmes d'information plus anciens exportent le plus simplement leurs données sous forme de texte brut.

Les documents en format de texte brut proposés pour l'archivage sous cette même forme peuvent être archivés sans autres dans ce format pour autant que le codage des caractères soit connu. Les codages de caractères courants ne présentent aucun problème, mais ils doivent être documentés dans les métadonnées. S'il existe une possibilité de choix, il faudrait utiliser de préférence UNICODE dans le codage UTF-8 en raison de sa compréhensibilité universelle et de sa durabilité.

Références

Le texte brut n'est pas une norme au sens formel. Il existe toutefois des normes formelles pour le codage des caractères .

Le type MIME text/plain dispose d'une définition formelle:

Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types, 1996

↗ <http://www.rfc-editor.org/rfc/rfc2046.txt>

Articles connexes

CSV

ODF

Informations générales

Titre	Open Document Format
Catégorie	Formats textuels
Abréviation	ODF
Extension de fichier	.odt, .odp
Mime Type	application/vnd.oasis.opendocument.text, application/vnd.oasis.opendocument.presentation
Pronom PUID	fmt/135; fmt/136; fmt/138; fmt/290; fmt/291; fmt/292; fmt/293 Version actuelle: 1.2 (2011 / OASIS et ISO/IEC 26300)
Version	Versions précédentes: 1.0 (2005 / OASIS et ISO/IEC 26300); 1.1 (2006 / OASIS) Version 1.2 en préparation

Description

ODF est un format fondé sur celui des fichiers du progiciel de bureautique source libre OpenOffice.org et donc sur XML. Un document ODF est un dossier ZIP ayant la structure d'une archive Java. Il comporte donc un fichier «Manifest» avec métadonnées ainsi que plusieurs autres fichiers dans lesquels se trouvent le document, des objets intégrés et d'autres métadonnées. Le contenu, la structure et la présentation du document sont enregistrés sous forme de fichiers XML. Les débats contradictoires sur ODF se réfèrent toujours à [OOXML](#). L'évaluation en tient occasionnellement compte.

Evaluation

Ouverture du format : 4

Ce format est une norme OASIS depuis 2005 et il est disponible publiquement depuis 2006 comme ISO/IEC 26300. L'algorithme de compression ZIP est également ouvert et libre d'accès. (<http://www.info-zip.org/>)

Licence libre : 3

Le format contient des brevets. Ceux-ci sont mis en circulation pour toutes les versions futures, tant que durera la collaboration avec Sun (ou Oracle).

Les polices utilisées ne font pas partie du domaine public, ce qui risque de poser des problèmes de droits lors de l'utilisation future.

Diffusion : 2

Ce format est encore peu répandu; il est cependant recommandé dans le domaine public ou prescrit généralement en tant que norme. (Allemagne, Belgique, Argentine, etc.).

Fonctionnalités : 4

ODF peut représenter toutes les propriétés significatives imaginables. OpenOffice permet de créer des documents ODF et peut convertir en ODF divers formats de traitement de texte courants et obsolètes. Préserver la présentation des documents ODF pose cependant problème. De même, les documents ODF originaux sont présentés sous diverses formes par les différents visualiseurs. Cela est dû au fait que les polices ainsi que les espaces de couleurs ne sont pas encapsulés dans les fichiers ODF. Les fichiers ODF peuvent contenir des macros, ce qui peut également empêcher un affichage fiable des documents. L'option de protection en écriture pose également des problèmes du point de vue archivistique.

Implémentation : 4

ODF est le format natif de l'application open source OpenOffice et de ses dérivés. Microsoft Office ainsi que d'autres produits permettent également de produire des documents à l'origine en ODF.

Densité de mémorisation : 3

La compression par ZIP permet d'atteindre une densité de mémorisation relativement élevée.

Vérifiabilité : 3

La reconnaissance peut se faire grâce à l'extension ainsi qu'à la séquence d'octets correspondante. Il existe un validateur ODF.

Bonnes pratiques : 1

L'utilisation de ce format par les autorités et dans l'administration est encouragée par l'Etat dans divers pays. Le format n'est pas encore utilisé par les services d'archivage.

Perspectives : 2

Les avis divergent quant aux perspectives de diffusion d'ODF. Quelques auteurs doutent que ce format puisse s'imposer malgré le soutien de l'Etat, compte tenu de la concurrence d' OOXML et de la position dominante de Microsoft sur le marché. Cela étant, une importante communauté largement soutenue travaille sur ce format. Il n'est par conséquent pas possible actuellement d'évaluer les chances d'avenir du format avec certitude.

Classe de formats : C

Ce format est intéressant pour l'archivage. Il n'est toujours pas répandu, bien qu'il existe depuis 2005.

Conclusion

L'affichage précis des documents ODF ne peut pas être reproduit. ODT est un format de données acceptable pour Office, mais il ne convient pas comme format d'archivage pour les données textuelles, en particulier à cause des éléments non encapsulés et des macros; les archives électroniques ne doivent contenir que des données sans aucun code source.

Références

ISO/IEC 26300-1:2015: Information technology — Open Document Format for Office Applications (OpenDocument) v1.2

↗ <https://www.iso.org/standard/66363.html>

[payant]

↗ <http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>

[téléchargeable gratuitement]

OASIS: OpenDocument 1.2

↗ <http://docs.oasis-open.org/office/v1.2/OpenDocument-v1.2.pdf>

ISO/IEC 29166:2011: Document description and processing languages — Guidelines for translation between ISO/IEC 26300 and ISO/IEC 29500 document formats (ODF et OOXML)

↗ <https://www.iso.org/standard/45245.html>

[payant]

Bibliographie

Wikipédia

↗ <http://fr.wikipedia.org/wiki/OpenDocument>

OpenOffice:

↗ <http://fr.openoffice.org/>

Articles connexes

OOXML

Catalogue des formats de données d'archivage

version 6.0, juil. 2019

Contact

A propos

Impressum

Événements

Newsletter

RSS

OOXML

Informations générales

Titre	Office Open Extensible Markup Language
Catégorie	Formats textuels
Abréviation	OOXML
Extension de fichier	.docx, .docm, .pptx, .pptm, .ppsx, .ppsm
Mime Type	application/vnd.openxmlformats-officedocument.wordprocessingml.document; application/vnd.ms-word.document.macroEnabled.12; application/vnd.openxmlformats-officedocument.presentationml.presentation; application/vnd.ms-powerpoint.presentation.macroEnabled.12; application/vnd.openxmlformats-officedocument.presentationml.slideshow; application/vnd.ms-powerpoint.slideshow.macroEnabled.12
Pronom PUID	fmt/189; fmt/215; fmt/412; fmt/487; fmt/523; fmt/629; fmt/630
Version	Version actuelle: 5 (2015/16 / ECMA 376 et ISO/IEC 29500:2016) Versions précédentes: 1 (2006 / ECMA 376); 2 (2008 / ECMA 376); 3 (2011 / ECMA 376 et ISO/IEC 29500:2012); 4 (2012 / ECMA 376)

Description

OOXML est un format de bureautique conçu par Microsoft sur la base de XML. Un format container est spécifié ainsi que des langages de balisage (markup language) basés sur XML pour les différentes applications bureautiques (traitement de texte, tableur, etc.). Un document OOXML est un fichier ZIP contenant différents documents XML qui eux-mêmes contiennent le contenu et la structure du document.

Evaluation

Ouverture du format : 4

OOXML est disponible publiquement; c'est une norme Ecma International (Ecma-376) depuis 2006 ainsi qu'une norme ISO/IEC (actuellement ISO/IEC 29500-1:2016) publiée depuis 2012. L'algorithme de compression ZIP est également ouvert et libre d'accès ([↗ http://www.info-zip.org/](http://www.info-zip.org/)).

Licence libre : 3

Office Open XML est disponible gratuitement et peut être copié. Le format contient cependant des brevets qui sont mis en circulation uniquement pour la version actuelle. Les polices utilisées ne font pas l'objet d'une licence libre, ce qui pourra causer des problèmes de droits lors de l'utilisation future des polices ainsi protégées.

Diffusion : 4

Comme c'est un format de fichiers de MS Office, OOXML est très répandu.

Fonctionnalités : 4

OOXML peut représenter toutes les propriétés significatives de documents textuels. Les mêmes objections que celles relatives au format ODF sont cependant formulées pour le format OOXML: la préservation de la présentation lors de conversions est médiocre, car les polices et les espaces de couleurs ne sont pas encapsulés dans OOXML. Les fichiers peuvent contenir des macros, ce qui peut également empêcher un affichage fiable des documents. En outre, l'option de protection en écriture disponible pose des problèmes du point de vue archivistique

Implémentation : 1

Outre l'implémentation d'origine dans MS Office, il existe d'autres implémentations indépendantes qui sont toutefois vraisemblablement incomplètes.

Densité de mémorisation : 3

L'utilisation de la compression ZIP permet d'atteindre une densité de mémorisation relativement élevée.

Vérifiabilité : 2

La reconnaissance est possible uniquement au moyen de l'extension. Il n'existe aucun validateur spécial.

Bonnes pratiques : 1

Ce format continue d'être considéré dans le monde de l'archivage comme un format de travail et non comme un format d'archivage.

Perspectives : 2

En raison de la large diffusion du format dans les administrations et pour éviter des migrations, il est envisageable de l'utiliser comme format d'archivage.

Classe de formats : B

Le format est utilisé dans les versions actuelles de MS Office.

Conclusion

A l'origine, OOXML est un format de données textuelles propriétaire pour lequel l'affichage des documents ne peut pas être reproduit de manière précise dans tous les cas. Les autres produits que ceux de Microsoft ne garantissent pas une implémentation correcte en raison de la complexité de la spécification. Le format s'est imposé comme format d'échange avec Office. Les éléments non encapsulés et les macros constituent un argument contre son utilisation comme format d'archivage étant donné que les archives électroniques ne devraient contenir que des données sans aucun code source. Cependant, la diffusion importante en tant que format de production fait qu'il sera probablement proposé ultérieurement aux archives.

Références

Norme ECMA-376, Office Open XML File Formats, 1st edition (décembre 2006), 2nd edition (décembre 2008), 3rd edition (juin 2011), 4th edition (December 2012) and 5th edition (Part 3, December 2015; and Parts 1&4, December 2016):

[↗ http://www.ecma-international.org/publications/standards/Ecma-376.htm](http://www.ecma-international.org/publications/standards/Ecma-376.htm)

Norme ISO/IEC 29500-1:2016

[↗ https://www.iso.org/standard/71691.html](https://www.iso.org/standard/71691.html)

[↗ http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html](http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html)

Microsoft Open Specification Promise:

[↗ https://msdn.microsoft.com/en-us/openspecifications](https://msdn.microsoft.com/en-us/openspecifications)

Bibliographie

Wikipédia: Office Open XML

[↗ http://fr.wikipedia.org/wiki/Office_Open_XML](http://fr.wikipedia.org/wiki/Office_Open_XML)

Ecma White Paper, Office Open XML Overview:

[↗ http://www.ecma-international.org/news/TC45_current_work/OpenXML%20White%20Paper.pdf](http://www.ecma-international.org/news/TC45_current_work/OpenXML%20White%20Paper.pdf)

Microsoft (ed.)

Le format Microsoft Office Open XML

[↗ http://msdn.microsoft.com/fr-ch/office/bb906068.aspx](http://msdn.microsoft.com/fr-ch/office/bb906068.aspx)

Articles connexes

[ODF](#)

Catalogue des formats de données d'archivage

version 6.0, juil. 2019

Contact
A propos
Impressum
Événements
Newsletter
RSS

[Formats de données \(Cfa\)](#) > [Données textuelles](#) > [PDF](#)

PDF

Informations générales

Titre	Portable Document Format
Catégorie	Formats textuels
Abréviation	PDF
Extension de fichier	.pdf
Mime Type	application/pdf
Pronom PUID	fmt/14; fmt/15; fmt/16; fmt/17; fmt/18; fmt/19; fmt/20; fmt/276 Version actuelle: 2 (2017), ou ISO 32000-2:2017
Version	Versions précédentes: 1.0 (1993), 1.1 (1994), 1.2 (1996), 1.3 (1999), 1.4 (2001), 1.5 (2003), 1.6 (2005), 1.7 (2006, ou ISO 32000:2008)

Description

Le «Portable Document Format» (PDF) est un format de fichier pour documents indépendant de la plate-forme. Le format de mémorisation et d'échange PDF est très répandu et apprécié. Il repose sur un langage de description de page à base vectorielle (développement basé sur PostScript) assurant l'évolutivité de la présentation.

Evaluation

Ouverture du format : 4

Le Portable Document Format (PDF) a été conçu par Adobe Systems et publié au moyen du PDF Reference Manual. PDF 1.7 a été standardisé en juillet 2008 en tant que norme ISO 32000-1. Les versions futures de PDF ne seront plus élaborées par Adobe, mais au moyen du processus de standardisation ISO; dernière version en date 2.0 (ISO 32000-2:2017)

Licence libre : 3

PDF ne fait pas l'objet d'une licence. La question de la licence pour les polices ou les algorithmes de compression utilisés doit être clarifiée au cas par cas.

Diffusion : 4

PDF existe depuis 1993 et est très répandu. La version 2 actuelle est disponible depuis juillet 2017. Il existe de nombreux outils sur diverses plateformes pour la conversion du format texte en PDF.

Fonctionnalités : 3

Tous les formats usuels de cette catégorie peuvent être convertis en PDF sans perte. Le contenu des fichiers peut faire l'objet d'une recherche intégrale du texte.

La mise en page peut être identique à l'original à condition de s'assurer que les polices et les espaces de couleur soient disponibles (par exemple en les encapsulant lors de la génération du fichier PDF).

Outre la structure, les documents PDF peuvent comporter des éléments interactifs comme des signets, des commentaires ou des champs de formulaire, ainsi que des objets audio/vidéo, de la transparence, des niveaux, des JavaScripts et des fichiers intégrés.

Des passages de texte, des tableaux et des graphiques (ou des parties de ceux-ci) provenant de documents PDF peuvent être collés dans d'autres programmes d'application pour y être traités.

Implémentation : 4

Il existe un grand nombre d'outils permettant de convertir différents formats en PDF. Citons PDFlib, PDF-Writer, Acrobat Distiller, MS Office 2016. Il existe de nombreuses visionneuses de PDF, parmi lesquelles également des produits open source.

Densité de mémorisation : 3

L'encombrement des documents PDF est en général relativement modeste. La compression permet de réduire la taille des fichiers PDF; consulter toutefois les considérations sur la compression des données .

Vérifiabilité : 4

La reconnaissance peut se faire grâce à l'extension ainsi qu'à la séquence d'octets correspondante. Les dernières versions comprennent des validateurs spéciaux.

Bonnes pratiques : 2

PDF est un format d'enregistrement et d'échange de données apprécié, très utilisé dans les administrations. Les documents sont convertis dans ce format dès qu'ils se trouvent sous leur forme définitive et ne seront plus modifiés. Les versions actuelles de PDF sont généralement utilisées à cet effet. Elles peuvent disposer de fonctionnalités comme le Web2.0 ou JavaScript et le chiffrement, qui ne sont pas appropriées dans le cas d'un archivage à long terme.

Les archives prennent en charge peu de PDF, car elles préfèrent les PDF/A .

Perspectives : 1

Il est prévisible que la popularité et la simplicité accrue de PDF sera accompagnée d'un développement continu. Les versions récentes de PDF comportent d'innombrables extensions faisant l'objet d'un brevet ou d'une licence et des fonctionnalités posant des problèmes pour l'archivage.

Classe de formats : B

Le format couvre les besoins d'utilisation globaux.

Conclusion

PDF est un format qui se prête en principe à l'archivage. Il faut toutefois prendre garde aux problèmes potentiels dus aux fonctionnalités spéciales et aux objets encapsulés. C'est pourquoi il est préférable de recourir si possible à PDF/A pour l'archivage. Dans le cas des documents dont la mise en page est importante pour l'archivage, il faut donc veiller à ce que tout le contenu, en particulier les images, les polices et les espaces de couleur soient encapsulés. La prudence est de mise lors du recours aux fonctions problématiques comme le Web2.0 ou JavaScript, les données intégrées et la production de sons ou de vidéos.

Références

ISO 32000-2:2017

Document management — Portable document format — Part 2: PDF 2

↗ <https://www.iso.org/standard/63534.html>

[payant]

Adobe (ed.)

PDF Reference

↗ http://www.adobe.com/devnet/pdf/pdf_reference.html

[gratuit]

Bibliographie

À propos d'Adobe PDF

↗ <https://acrobat.adobe.com/ch/fr/why-adobe/about-adobe-pdf.html>

Wikipédia: Portable Document Format

↗ http://fr.wikipedia.org/wiki/Portable_Document_Format

PDF Tools AG (ed)

White Paper: PDF. Les bases

↗ <http://www.pdf-tools.com/public/downloads/whitepapers/Whitepaper-PDF-Grundlagen-DE.pdf>

Parker, Nick; Bouloc, Michel; Rouchon, Olivier

Guide méthodologique. Le format de fichiers PDF

2012

↗ <http://www.huma-num.fr/ressourcesdoc/guide-format-pdf-1>

Articles connexes

[PDF/A-1](#) , [PDF/A-2](#) , [PDF/A-3](#)

Catalogue des formats de données d'archivage

version 6.0, juil. 2019

Contact

A propos

Impressum

Événements

Newsletter

RSS

[Formats de données \(Cfa\)](#) > [Données textuelles](#) > [PDF/A-1](#)

PDF/A-1

Informations générales

Titre	Portable Document Format (Archival)
Catégorie	Formats textuels
Abréviation	PDF/A
Extension de fichier	.pdf
Mime Type	application/pdf
Pronom PUID	fmt/95; fmt/354
Version	PDF/A-1 (correspond à la version 1.4 PDF de 2001) avec PDF/A 1a et PDF/A 1b (2005) <u>PDF/A-2</u> et <u>PDF/A-3</u> sont également publiés. Comme aucune des versions PDF/A ne remplace une autre, toutes les versions sont valables au même titre.

Description

PDF/A-1 est un «Portable Document Format» conçu pour l'archivage. Le format a été spécifié dans la norme «ISO-19005-1 – Document management – Electronic document file format for long-term preservation». Cette norme correspond à la version PDF 1.4. La norme ne donne que la liste des fonctions des versions PDF qui sont obligatoires, recommandées, restreintes ou interdites.

On fait la distinction entre

PDF/A 1a: conformité intégrale à la norme PDF/A

PDF/A 1b: satisfait aux exigences minimales de PDF/A (l'accessibilité, conformément au paragraphe 508 de l'acte juridique «US Rehabilitation Act», fait défaut).

Evaluation

Ouverture du format : 4

PDF/A-1 est fondé sur la version 1.4 du cahier des charges PDF , élaboré et mis à disposition par l'entreprise Adobe Systems. PDF/A-1 est publié en tant que standard ISO 19005-1 et donc ouvert.

Licence libre : 3

Le critère de licence libre est rempli, en particulier car il n'est pas possible d'utiliser tous les algorithmes de compression.

Diffusion : 4

PDF/A-1 correspond à la version PDF 1.4 (avec certaines restrictions). L'utilisation de cette version est par conséquent en diminution. PDF/A-1 lui-même est cependant répandu et connu. Sa diffusion va continuer à gagner du terrain, notamment du fait que la version 1 demeure valable et qu'elle ne doit plus migrer.

Fonctionnalités : 3

Contrairement à PDF 1.4, la transparence, la production de sons/vidéos et les objets JavaScript ne sont pas autorisés. Les objets structurés (textes, graphiques vectoriels, graphiques matriciels) sont conservés pour autant qu'ils soient encapsulés. WYSIWYG est assuré moyennant quelques petites restrictions. La reproduction visuelle et l'ordre naturel de lecture sont garantis.

Dans le cas de PDF/A-1a, l'extraction de texte et la facilité de recherche sont préservées. Dans le cas de PDF/A-1b, elles ne sont pas garanties.

Implémentation : 4

Il existe différents outils pour produire des documents PDF/A parmi lesquels figurent également des produits open source. Les fichiers PDF/A peuvent être lus avec tous les lecteurs de PDF (par ex. Adobe Reader).

Densité de mémorisation : 2

PDF/A est en règle générale relativement compact. L'encapsulation des polices nécessite toutefois davantage de mémoire que dans PDF .

Vérifiabilité : 4

Les fichiers PDF/A-1 peuvent être reconnus au moyen de la séquence d'octets. Il existe de nombreux validateurs indépendants les uns des autres.

Bonnes pratiques : 4

PDF/A-1b est un format d'archivage déjà utilisé et exigé par de nombreux services. PDF/A-1 convient particulièrement à l'archivage à long terme et peut, en outre, être imprimé sans perte d'informations.

Perspectives : 3

La norme PDF/A-1b s'est imposée pour l'archivage à long terme. Etant donné que la technologie utilisée date un peu, PDF/A-2 permet toutefois d'obtenir de meilleurs résultats lors de la création de documents. Des nouvelles versions sont développées, mais elles n'annulent pas cette version.

Classe de formats : B

PDF/A a été conçu spécifiquement pour l'archivage et ne va pas à l'encontre des exigences relatives à l'archivage à long terme.

Conclusion

PDF/A-1 est un format très approprié pour l'archivage à long terme.

Le CECO recommande la version PDF/A-2 au niveau de conformité U pour la création de nouveaux documents PDF/A afin de permettre la transparence et de garantir la facilité de recherche. Les fichiers PDF/A-1 existants ne présentent pas de risque pour l'archivage à long terme et ne doivent en aucun cas être actualisés.

Planification de la pérennisation

La méthode de compression JBIG2 autorisée dans le PDF/A sauvegarde une seule fois des symboles d'égale valeur et les utilise à plusieurs reprises dans tout le document. Ce procédé appelé pattern matching and substitution (PMS) a pour conséquence, lors de certaines implémentations et configurations, de remplacer des signes par d'autres dans les textes. Comme l'erreur est irréversible et qu'il est impossible d'établir si le procédé PMS erroné a été utilisé ou pas, le CECO recommande de renoncer pour le moment à la compression JBIG2 lors de la création de fichiers PDF. Pour plus de détails, voir [↗ https://kost-ceco.ch/cms/index.php?jbig2-compression_fr](https://kost-ceco.ch/cms/index.php?jbig2-compression_fr) .

Références

ISO 19005-1:2005, Gestion de documents — Format de fichier des documents électroniques pour une conservation à long terme — Partie 1: Utilisation du PDF 1.4 (PDF/A-1)

↗ <https://www.iso.org/standard/38920.html>

[payant]

PDF_Reference_1.4

PDF Reference (third edition) Adobe Portable Document Format Version 1.4

↗ http://www.adobe.com/devnet/pdf/pdf_reference_archive.html

[gratuit]

Bibliographie

PDF Tools AG (ed)

White Paper: PDF/A - der Standard für die Langzeitarchivierung

↗ <https://www.pdf-tools.com/public/downloads/know-how/whitepaper-pdf-a-standard-iso-19005-de.pdf>

PDF Association

↗ <https://www.pdfa.org/>

Articles connexes

[PDF](#) , [PDF/A-2](#) , [PDF/A-3](#)

Catalogue des formats de données d'archivage

version 6.0, juil. 2019

Contact

A propos

Impressum

Événements

Newsletter

RSS

PDF/A-2

Informations générales

Titre	Portable Document Format (Archival)
Catégorie	Formats textuels
Abréviation	PDF/A, PDF/A-2, PDF/A-2a, PDF/A-2b, PDF/A-2u
Extension de fichier	.pdf
Mime Type	application/pdf
Pronom PUID	fmt/476, fmt/477, fmt/478 PDF/A-2 (correspond à ISO-32000-1 ou à la version PDF 1.7) avec PDF/A-2a, PDF/A-2b et PDF/A-2u (2011)
Version	PDF/A-1 et PDF/A-3 sont également publiés. Comme aucune des versions PDF/A ne remplace une autre, toutes les versions sont valables au même titre.

Description

PDF/A-2 est un «Portable Document Format» conçu pour l'archivage. Le format a été spécifié dans la norme «ISO-19005-2 – Document management – Electronic document file format for long-term preservation». Cette norme correspond à la version [PDF](#) ISO 32000-1. La norme ne donne que la liste des fonctions des versions PDF à la base qui sont obligatoires, recommandées, restreintes ou interdites.

On fait la distinction entre

PDF/A 2a: conformité intégrale à la norme PDF/A;

PDF/A 2b: satisfait aux exigences minimales de PDF/A (l'accessibilité, conformément au paragraphe 508 de l'acte juridique «US Rehabilitation Act», fait défaut);

PDF/A-2u: comme PDF/A-2b, mais en plus, l'ensemble du texte est reproduit en Unicode afin qu'il puisse être indexé et affiché

Evaluation

Ouverture du format : 4

PDF/A-2 est fondé sur la version ISO-32000-1 du cahier des charges PDF, qui repose elle-même sur la version 1.7 élaborée par l'entreprise Adobe Systems. PDF/A-2 est publié en tant que norme ISO 19005-2 et donc ouvert.

Licence libre : 3

Le critère de licence libre est rempli, en particulier car il n'est pas possible d'utiliser tous les algorithmes de compression.

Diffusion : 4

Dans la production, PDF/A-2 à entre-temps en grande partie remplacé PDF/A-1.

Fonctionnalités : 3

Contrairement à la version PDF ISO-32000-1, la production de son/vidéo et les objets JavaScript ne sont pas autorisés. Les objets structurés (textes, graphiques vectoriels, graphiques matriciels) sont conservés pour autant qu'ils soient encapsulés.

La reproduction visuelle et l'ordre naturel de lecture sont garantis.

WYSIWYG est assuré moyennant quelques petites restrictions.

Implémentation : 4

Il existe différents outils permettant de créer des documents PDF/A, parmi eux également des produits open source. Les fichiers PDF/A peuvent être lus avec tous les lecteurs de PDF (p. ex. Adobe Reader).

Densité de mémorisation : 3

PDF/A-2 est en règle générale compact. Au contraire de PDF/A-1, la compression JPEG2000 est autorisée et la densité de mémorisation peut ainsi aussi être plus élevée. L'encapsulation des polices nécessite toutefois davantage de mémoire que dans PDF.

Vérifiabilité : 4

Les fichiers PDF/A-2 peuvent être reconnus au moyen de la séquence d'octets. Il existe de nombreux validateurs.

Bonnes pratiques : 4

PDF/A-2 est accepté et utilisé comme format d'archivage par plusieurs services.

PDF/A-2 convient particulièrement à l'archivage à long terme, mais il ne peut pas être imprimé sans perte d'informations.

Perspectives : 4

La norme PDF/A-2b s'est imposée pour l'archivage à long terme aux côtés de PDF/A-1b.

De nouvelles versions sont développées, mais elles n'annulent pas cette version.

Classe de formats : B

Le format est actuellement en usage.

Conclusion

PDF/A-2 est un format très approprié pour l'archivage à long terme.

Les services d'archivage devraient encourager le stockage des documents administratifs dans la norme PDF/A-2u. Les bonnes pratiques en matière de PDF/A montrent que non seulement la reproduction visuelle, mais également la facilité de recherche et l'extraction sont des conditions, raisons pour lesquelles le niveau de conformité U devrait être préféré au B. Si des documents sont enregistrés sous forme de portfolio PDF/A-2, il faut veiller à ne pas utiliser une structure de contenu importante.

Planification de la pérennisation

La méthode de compression JBIG2 autorisée dans le PDF/A sauvegarde une seule fois des symboles d'égale valeur et les utilise à plusieurs reprises dans tout le document. Ce procédé appelé pattern matching and substitution (PMS) a pour conséquence, lors de certaines implémentations et configurations, de remplacer des signes par d'autres dans les textes. Comme l'erreur est irréversible et qu'il est impossible d'établir si le procédé PMS erroné a été utilisé ou pas, le CECO recommande de renoncer pour le moment à la compression JBIG2 lors de la création de fichiers PDF. Pour plus de détails, voir https://kost-ceco.ch/cms/index.php?jbig2-compression_fr.

Références

ISO 19005-2:2011

Document management — Electronic document file format for long-term preservation — Part 2: Use of ISO 32000-1 (PDF/A-2)

<https://www.iso.org/standard/50655.html>

[payant]

ISO 32000-1:2008

Document management — Portable document format — Part 1: PDF 1.7

<http://www.iso.org/standard/51502.html>

[payant]

Bibliographie

CECO

PDF/A-2 et PDF/A-3: quelles nouveautés?

Etude du CECO, 2013

https://kost-ceco.ch/cms/index.php?pdf-a-2_3_study_fr

PDF Tools AG (ed)

White Paper: PDF/A - der Standard für die Langzeitarchivierung

<https://www.pdf-tools.com/public/downloads/know-how/whitepaper-pdf-a-standard-iso-19005-de.pdf>

PDF Association

<https://www.pdfa.org/>

Articles connexes

[PDF](#) , [PDF/A-1](#) , [PDF/A-3](#)

Catalogue des formats de données d'archivage

version 6.0, juil. 2019

Contact
A propos
Impressum
Événements
Newsletter
RSS

PDF/A-3

Informations générales

Titre	Portable Document Format (Archival)
Catégorie	Formats textuels
Abréviation	PDF/A, PDF/A-3, PDF/A-3a, PDF/A-3b, PDF/A-3u
Extension de fichier	.pdf
Mime Type	application/pdf
Pronom PUID	fmt/479, fmt/480, fmt/481
Version	PDF/A-3 (correspond à ISO-32000-1 ou à la version PDF 1.7) avec PDF/A-3a, PDF/A-3b et PDF/A-3u (2012) – PDF/A-1 et PDF/A-2 sont également publiés. Comme aucune des versions PDF/A ne remplace une autre, toutes les versions sont valables au même titre.

Description

PDF/A-3 est un «Portable Document Format» conçu pour l'archivage. Le format a été spécifié dans la norme «ISO-19005-3 – Document management – Electronic document file format for long-term preservation». Cette norme correspond à la version PDF ISO 32000-1. La norme ne donne que la liste des fonctions de chacune des versions PDF qui sont obligatoires, recommandées, restreintes ou interdites.

On fait la distinction entre

PDF/A 3a: conformité intégrale à la norme PDF/A

PDF/A 3b: satisfait aux exigences minimales de PDF/A (l'accessibilité, conformément au paragraphe 508 de l'acte juridique «US Rehabilitation Act», fait défaut);

PDF/A-3u: comme PDF/A-3b, mais en plus l'ensemble du texte est reproduit en Unicode afin qu'il puisse être indexé et affiché.

Evaluation

Ouverture du format : 4

PDF/A-3 repose sur la version ISO-32000-1 du cahier des charges PDF, qui repose elle-même sur la version 1.7 élaborée par l'entreprise Adobe Systems. PDF/A-3 est publié en tant que norme ISO 19005-3 et donc ouvert.

Licence libre : 1

Le critère de licence libre est rempli uniquement pour le fichier PDF/A-3, en particulier car il n'est pas possible d'utiliser tous les algorithmes de compression. Les fichiers encapsulés peuvent cependant comporter des licences.

Diffusion : 1

PDF/A-3 correspond à la version PDF ISO-32000-1 (avec certaines restrictions). Cette version est de plus en plus utilisée. PDF/A-3 lui-même est encore peu répandu et peu connu.

Fonctionnalités : 3

A une exception près, les fonctionnalités sont identiques à celles de PDF/A-2.

Au contraire de PDF/A-2, tous les fichiers encapsulés sont autorisés.

Implémentation : 3

Il existe différents outils permettant de créer des documents PDF/A, parmi lesquels figurent également des produits open source. Les fichiers PDF/A peuvent être lus avec tous les lecteurs de PDF (par ex. Adobe Reader). Il existe cependant très peu de lecteurs permettant de lire la structure d'un portfolio PDF.

Densité de mémorisation : 2

PDF/A-3 est en règle générale compact. Au contraire de PDF/A-2, les fichiers encapsulés de tout type nécessitent cependant davantage de mémoire.

Vérifiabilité : 2

Les fichiers PDF/A-3 peuvent être reconnus au moyen de la séquence d'octets. Il existe de nombreux validateurs. Les fichiers encapsulés ne peuvent cependant pas être reconnus ni validés.

Bonnes pratiques : 1

PDF/A-3 n'est encore que peu utilisé, mais il est accepté comme format d'archivage par de nombreuses archives.

PDF/A-3 ne convient pas à l'archivage à long terme ni aux archives publiques. Il peut cependant être intéressant pour d'autres archives, comme p. ex. les archives d'entreprises pour lesquelles les durées d'archivage sont limitées.

Perspectives : 1

La norme PDF/A-3 ne s'imposera très vraisemblablement pas pour l'archivage à long terme.

De nouvelles versions sont développées, mais elles n'annulent pas cette version.

Classe de formats : Ø

PDF/A-3 a été conçu spécifiquement pour l'archivage, mais ne répond pas aux exigences relatives à l'archivage à long terme.

Conclusion

PDF/A-3 n'est pas un format adapté à l'archivage à long terme.

Les Archives devraient encourager le stockage des documents administratifs dans la norme PDF/A-1 ou PDF/A-2. Si des documents sont enregistrés sous forme de portfolio PDF/A-2, il faut veiller à ne pas utiliser une structure de contenu importante. Les bonnes pratiques montrent clairement que pour le PDF/A, non seulement la reproduction visuelle, mais également la facilité de recherche et l'extraction sont des conditions, raisons pour lesquelles le niveau de conformité U devrait être préféré au B.

Planification de la pérennisation

La méthode de compression JBIG2 autorisée dans le PDF/A sauvegarde une seule fois des symboles d'égale valeur et les utilise à plusieurs reprises dans tout le document. Ce procédé appelé pattern matching and substitution (PMS) a pour conséquence, lors de certaines implémentations et configurations, de remplacer des signes par d'autres dans les textes. Comme l'erreur est irréversible et qu'il est impossible d'établir si le procédé PMS erroné a été utilisé ou pas, le CECO recommande de renoncer pour le moment à la compression JBIG2 lors de la création de fichiers PDF. Pour plus de détails, voir https://kost-ceco.ch/cms/index.php?jbig2-compression_fr.

Références

ISO 19005-3:2012

Document management — Electronic document file format for long-term preservation — Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)

<https://www.iso.org/standard/57229.html>

[payant]

ISO 32000-1:2008

Document management — Portable document format — Part 1: PDF 1.7

[↗ https://www.iso.org/standard/51502.html](https://www.iso.org/standard/51502.html)

[payant]

Bibliographie

CECO

PDF/A-2 et PDF/A-3: quelles nouveautés?

Etude du CECO, 2013

[↗ https://kost-ceco.ch/cms/index.php?pdf-a-2_3_study_fr](https://kost-ceco.ch/cms/index.php?pdf-a-2_3_study_fr)

PDF Tools AG (ed)

White Paper: PDF/A - der Standard für die Langzeitarchivierung

[↗ https://www.pdf-tools.com/public/downloads/know-how/whitepaper-pdf-a-standard-iso-19005-de.pdf](https://www.pdf-tools.com/public/downloads/know-how/whitepaper-pdf-a-standard-iso-19005-de.pdf)

PDF Association

[↗ https://www.pdfa.org/](https://www.pdfa.org/)

Articles connexes

[PDF](#) , [PDF/A-1](#) , [PDF/A-2](#)

Catalogue des formats de données d'archivage

version 6.0, juil. 2019

Contact
A propos
Impressum
Événements
Newsletter
RSS